

# DIE DAC-EVALUIERUNGSKRITERIEN: ZWISCHEN OPTIMIERUNG UND TRANSFORMATION

## Zusammenfassung

Kaum eine andere Harmonisierungsleistung hat in der internationalen Entwicklungszusammenarbeit (EZ) derart breite Anwendung erfahren wie die fünf Evaluierungskriterien des Entwicklungsausschusses der OECD (DAC), die 1991 verfasst wurden. Die Bewertung von Relevanz, Effektivität, Effizienz, Impact und Nachhaltigkeit einer entwicklungspolitischen Maßnahme leistet in der konkreten Anwendung einen wertvollen Strukturierungsbeitrag zur Evaluierung der EZ. Den Nutzen des Kriterienkatalogs kennzeichnen vier Funktionen:

- **Referenzfunktion:** die Bereitstellung eines zur inhaltlichen Neutralität verpflichteten Referenzrahmens, der als Grundlage einer unabhängigen Bewertung von EZ Maßnahmen herangezogen wird;
- **Anreizfunktion:** das Setzen von Anreizen, Maßnahmen entlang zentraler Kriterien entwicklungspolitischer Wirksamkeit zu konzipieren und durchzuführen;
- **Lernfunktion:** Maßnahmen entlang derselben Evaluierungskriterien zu vergleichen, eröffnet Lernpotenzial durch die Möglichkeit von Aggregation und Diffusion von Wissen;
- **Tiefenfunktion:** der Anspruch einer umfassenden Bewertung von Maßnahmen von der Relevanz einer Maßnahme bis hin zu deren Nachhaltigkeit.

Trotz dieses Erfolgs hat sich die Diskussion um die Angemessenheit und Nützlichkeit des Kriterienkatalogs in letzter Zeit intensiviert. Im Diskurs werden nicht nur immer wieder technische Aspekte der Optimierung genannt, die sich auf die Präzisierung und Trennschärfe der Kriterien beziehen. Es wird auch vermehrt die Frage gestellt, inwiefern der Kriterienkatalog grundsätzlich transformiert werden sollte. Thematisiert werden dabei u. a. eine normativ-inhaltliche Aufladung der Kriterien entlang aktueller globaler Agenden – wie etwa der Sustainable Development Goals (SDG) und der Agenda 2030 (UN 2015) - sowie teilweise auch die stärkere Betonung von Evaluierung als Instrument der

direkten Förderung von Entwicklungsprozessen. Zudem wird die Erweiterung des Kriterienkatalogs um Kriterien wie *Kohärenz* oder *Synergien* in Erwägung gezogen und diskutiert. Auch die Debatte, inwieweit die Anwendung des Kriterienkatalogs stärker standardisiert werden sollte, steht im Raum.

Vor diesem Hintergrund will der Beitrag die aktuelle Debatte strukturieren und erste Empfehlungen für deren Weiterentwicklung geben. Dabei stehen drei identifizierte Spannungsfelder im Mittelpunkt:

- das Spannungsfeld zwischen einer inhaltlichen Ausrichtung der Kriterien und deren ursprünglichem Anspruch auf inhaltliche Neutralität;
- das Spannungsfeld zwischen der operativen Handhabbarkeit des Katalogs sowie einer vollständigen Listung von evaluierungsrelevanten Aspekten;
- das Spannungsfeld zwischen Standardisierung und Flexibilisierung in der Anwendung des Katalogs.

Nach Auffassung der Autoren ist dabei eine Reform des Katalogs einer umfassenden Transformation vorzuziehen. Eine solche Reform sollte

- eine moderate inhaltliche Ausrichtung des Kriterienkatalogs entlang der Agenda 2030 beinhalten,
- die Anzahl der Kriterien im Sinne der operativen Handhabbarkeit des Kriterienkatalogs allenfalls moderat erhöhen,
- die Trennschärfe der Kriterien erhöhen und die Standardisierung in der Anwendung des Katalogs stärken, um organisationsübergreifende Lerneffekte zu verbessern.

### BOX 1: Zur Normierung von Evaluierung: Standards, Kriterien und Prüffragen

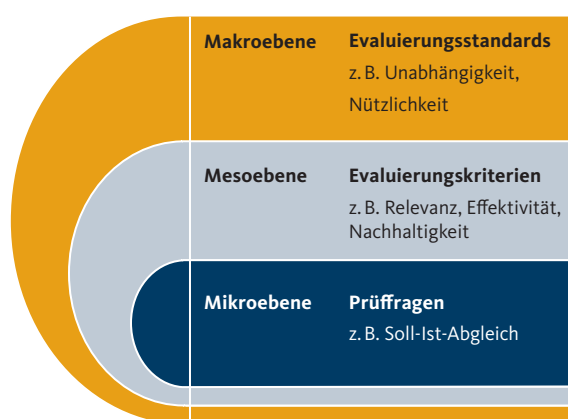
Innerhalb der EZ der OECD-Staaten werden die konkreten Fragestellungen und Methoden der Evaluierung in einen normierenden Rahmen eingebettet. Die Makroebene dieses Rahmenwerks steht für *Evaluierungsstandards*, die zentrale Aspekte „guter“ Evaluierung vorgeben, z. B. Unabhängigkeit, Nützlichkeit, Genauigkeit, Nachvollziehbarkeit oder ethische Verantwortung. Auf der Mesoebene ist der hier thematisierte Katalog mit *Evaluierungskriterien* angesiedelt, an dem sich eine Evaluierung konzeptionell ausrichten soll. Die Kriterien sollen strukturierend auf die inhaltliche Ausrichtung der Evaluierungsfragen wirken und die eingangs genannten Funktionen z. B. hinsichtlich Lernen erfüllen. Auf der Mikroebene werden die Evaluierungskriterien durch eine größere Anzahl an *Prüffragen* spezifiziert, deren Anwendung durch die einzelnen Entwicklungsorganisationen jedoch flexibel gehandhabt wird (siehe Box 2).

development“ – Evaluierung als Instrument, das über seine Lern- und Kontrollfunktion eher indirekt auf die Verbesserung der EZ wirkt, indem v. a. Planungs- und Umsetzungsstrukturen und -prozesse verbessert werden sollen.

Gleichwohl fordern Kritiker teilweise ein, dass der Kriterienkatalog sich direkter in den Dienst der SDGs stellen müsse und Evaluierung als unmittelbar auf Entwicklungsziele wirkendes Instrument einzusetzen sei. Eine solche Integration inhaltlicher Aspekte birgt jedoch Risiken. Dadurch würde das Bestreben nach höchstmöglicher Neutralität von Evaluierung aufgeweicht und indirekt die Glaubwürdigkeit beschädigt.

Zugleich eröffnen gerade die Prinzipien der Agenda 2030 neue Perspektiven, die inhaltlich wertneutral beurteilt werden können. In einen überarbeiteten Kriterienkatalog könnten etwa innovative Aspekte der Agenda 2030 eingeflochten werden, z. B. durch ein Mainstreaming in allen Kriterien über neue oder angepasste Prüffragen. Zudem könnten für die Agenda 2030 relevante Befunde einer Evaluierung zusammenfassend und gesondert dargestellt werden.

Abbildung zu Box 1



Quelle: eigene Darstellung

## 1. Neutralität versus inhaltliche Aufladung

Der 1991 formulierte Kriterienkatalog und die ihm zugrundeliegenden Evaluierungskriterien und -standards der OECD wurden bislang nicht inhaltlich aufgeladen, sondern sind von eher zeitlosem Charakter. Es wird weder explizit Bezug auf die politischen und ökonomischen Liberalisierungstrends der 1990er Jahre genommen, noch auf die Millenniumsziele der Vereinten Nationen. Darüber hinaus versteht der OECD-Kriterienkatalog – im Gegensatz zu direkteren Ansätzen wie etwa „evaluation for

## 2. Umfassende Betrachtung versus Machbarkeit

Der Kriterienkatalog schlägt nach OECD-DAC neben den fünf Kriterien auch 13 nachgeordnete Prüffragen vor (zur Begriffssystematik siehe Abbildung bzw. Box 1). Im Anschluss an die Einstiegsfrage nach den richtigen Zielen (*Relevanz*) soll analysiert werden, ob diese erreicht wurden (*Effektivität*), welche weiteren Veränderungen eingetreten sind (*Impact*), mit welchem Aufwand dies geschehen ist (*Effizienz*) und hinsichtlich welchen Zeithorizonts (*Nachhaltigkeit*). In der durch das BMZ (2006) adaptierten Version des Katalogs wird zudem ein gesondertes Kriterium zu *Koordination, Kohärenz und Komplementarität* bearbeitet sowie die Kriterien umfassender definiert.

In der internationalen Diskussion wird die Aufnahme weiterer Kriterien angeregt. Neben *Kohärenz* werden auch *Innovation* oder *Resilienz* als weitere, mögliche Kriterien genannt, die stärker auf aktuelle Anforderungen der EZ rekurrieren. Gleichwohl beeinträchtigt eine Vielzahl weiterer Kriterien eine fundierte Bewertung aufgrund der begrenzten Ressourcen. Der Überblick über die eher technischen Herausforderungen (Box 2) zeigt eine ganze Reihe an Unschärfen bei der Handhabung der Evaluierungskriterien.

## Box 2: Aktuelle Diskussion der Herausforderungen bei der Umsetzung und Konzeption der DAC-Kriterien

In der national und international geführten Debatte um den Anpassungsbedarf der DAC-Kriterien hat das DEval in einem Recherche- und Dialogprozess die gängigsten Argumente zusammengetragen.

Das **Relevanzkriterium** prüft, inwiefern Themen bzw. Ziele einer Maßnahme entwicklungspolitisch sinnvoll gewählt bzw. ob durchgeführte Aktivitäten auf die Zielerreichung gerichtet sind. In der praktischen Umsetzung wird die Vielzahl an nationalen wie internationalen Agenden und Zielen kritisiert. Sie erschwert eine Priorisierung entlang zentraler Defizite oder positiver Dynamiken. Aufgrund dieses Orientierungsdefizits findet oft eine überdurchschnittlich positive Bewertung der Relevanz einer Maßnahme statt.

Das **Effektivitätskriterium** sieht den Abgleich der geplanten mit den tatsächlich erreichten Zielen vor. In der Praxis werden jedoch häufig unterschiedliche Wirkungsebenen betrachtet, etwa die Ebene der unmittelbaren Leistungen einer Maßnahme oder die Ebene der Effekte auf die Zielgruppe. Ein weiterer Diskussionsstrang stellt heraus, dass bei der Evaluierung von Effektivität und Impact nur selten die verzerrende Wirkung weiterer Einflussgrößen auf die Zielgröße systematisch berücksichtigt werden.

Im Zentrum der **Impacteinschätzung** stehen beabsichtigte und nicht beabsichtigte, positive und negative übergeordnete Veränderungen durch eine Entwicklungsmaßnahme. Aus Sicht der Evaluierenden wird häufig auf die Herausforderung hingewiesen, Veränderungen auf übergeordneter Ebene konkreten Maßnahmen zuzuordnen. Meta-Evaluierungen zeigen ebenfalls, dass nicht-intendierte Veränderungen eher selten systematisch erfasst werden.

Die **Effizienzbewertung** setzt den Kostenaufwand einer Maßnahme zu deren Wirkungen in Bezug. Sie zielt in der Praxis meist auf die Identifikation von Optimierungspotenzial bei der Durchführung eines Maßnahmentyps. Die Diskussion regt an, zusätzlich Erkenntnisse über die relative Wirksamkeit unterschiedlicher entwicklungspolitischer Ansätze oder Instrumente zu generieren. Eine hierfür notwendige, vergleichend angelegte Bewertung findet jedoch aufgrund des sehr hohen Aufwands nur in Ausnahmefällen statt.

Das **Nachhaltigkeitskriterium** beinhaltet zum einen die Fortdauer erreichter Wirkungen nach Beendigung der Maßnahme, zum anderen die Idee der Generationengerechtigkeit zwischen Gegenwart und Zukunft (bzw. der Verschränkung von sozialen, ökonomischen und ökologischen Wirkungen). Die Verankerung dieser Elemente in nur einem Kriterium bewirkt allerdings häufig, dass nur einer der beiden Aspekte in der Evaluierung berücksichtigt wird oder dass beide Aspekte in unzureichender Tiefe behandelt werden.

In der Diskussion der DAC-Kriterien wird darüber hinaus auf eine ganzheitliche Betrachtung hingewiesen, demzufolge auch die **Kohärenz** zwischen Entwicklungsansätzen und -maßnahmen, sowie den hier diskutierten Kriterien mit einbezogen werden sollte. Zudem stehe die mechanische Anwendung der Kriterien einer Anpassung auf den spezifischen Kontext entgegen.

Eine Aufnahme weiterer Kriterien jenseits einer zusammenfassenden Betrachtung relevanter Bezüge zur Agenda 2030 birgt die Gefahr, die bereits bestehenden konzeptionellen Herausforderungen zu verstärken. Das Augenmerk einer Reform des Kriterienkatalogs sollte sich daher primär auf die Bearbeitung der existierenden Unschärfen und Überlappungen konzentrieren und versuchen, neue Aspekte auf der Mikroebene der Prüffragen zu integrieren. So ließe sich z. B. die Frage nach der Stärkung von *Resilienz* durch eine Maßnahme unter dem Kriterium der Nachhaltigkeit verorten, Aspekte von *Kohärenz* und *Synergien* könnten als Prüffragen unter dem Relevanz- und dem Impactkriterium bearbeitet werden. Eine definitorische Präzisierung der Kriterien fördert die kohärente Entwicklung von Prüffragen und die Vergleichbarkeit zwischen Kriterien.

### 3. Standardisierung versus Flexibilisierung

Im Rahmen der Entwicklung von Prüffragen stellt sich die Frage nach einem angemessenen Grad an Flexibilität in der praktischen Anwendung des Kriterienkatalogs. Laut BMZ (2006) steht es Anwendern des Katalogs im Rahmen der deutschen bilateralen EZ in begrenztem Rahmen offen, welche Prüffragen bearbeitet werden. Eine zusammenfassende Betrachtung der Ergebnisse zur Agenda 2030 würde eine vergleichende Berichterstattung aus Evaluierungen zu dieser Thematik begünstigen. Gleichzeitig bestünde innerhalb der Kriterien eine vergleichsweise flexible Ausrichtung auf spezifische Lern- und Kontrollbedürfnisse einzelner Evaluationsmaßnahmen fort.

Zugleich beschränkt die Flexibilisierung die Vergleichbarkeit von Evaluierungsergebnissen und hemmt interorganisationales Lernen aus Evaluierungen. Insgesamt sollte eine Reform des Kriterienkatalogs die Vorteile der Standardisierung bzw. des daraus resultierenden Lernpotenzials nicht vernachlässigen. Dies kann einerseits durch die bereits genannte Schärfung von Definitionen bei den Kriterien und Prüffragen erfolgen, sodass deren Bearbeitung in dieselbe Richtung geht. Andererseits bietet es sich an, bei den Prüffragen eine Priorisierung vorzunehmen, aus der zumindest eine Empfehlung hervorgeht, welche Fragen bei allen Evaluierungen vorrangig behandelt werden sollten. Auf Ebene der Kriterien sorgt also der Einsatz der hier diskutierten Kriterien für eine stärkere Vergleichbarkeit, während auf Ebene der Prüffragen Flexibilität ermöglicht wird.

## Fazit

Trotz seiner Errungenschaften bedarf der Kriterienkatalog zu Evaluierung von entwicklungspolitischen Maßnahmen einer Überarbeitung. Um die bestehenden Vorteile und Funktionen des Kataloges weiterhin nutzen zu können und zu stärken, sollte eine solche Überarbeitung eher den Charakter einer Reform

tragen und keine grundlegende Transformation anstreben. Hierbei plädieren wir für

- eine moderate inhaltliche Ausrichtung auf die Prinzipien der Agenda 2030 für Nachhaltige Entwicklung,
- eine Schärfung der bestehenden Kriterien und Prüffragen bei gleichzeitiger Ergänzung um relevanter gewordene Aspekte wie *Resilienz*, *Synergien* oder *Kohärenz* sowie
- Empfehlungen zur Priorisierung einzelner Prüffragen, insbesondere falls in der Überarbeitung neue Prüffragen entwickelt werden, um die Vergleichbarkeit von Evaluierungen auch über Organisationen hinweg zu stärken und dennoch Raum für Flexibilität zu lassen.

Eine solche Reform könnte die existierenden Vorteile des Kriterienkatalogs stärken und das übergreifende Lernen aus Evaluierungen fördern, ohne dass der Wiedererkennungswert und die Referenzfunktion des Kriterienkatalogs Schaden nehmen. Ein dergestalt reformierter Katalog böte zudem die Chance einer Anwendung jenseits der EZ und würde dem Universalitätsprinzip der Agenda 2030 gerecht.

## Literatur

**OECD DAC (1991)**, *DAC Criteria for Evaluating Development Assistance*, Paris.

**BMZ (2006)**, *Evaluierungskriterien für die deutsche bilaterale Entwicklungszusammenarbeit*, Bonn.

**UN (2015)**, *Resolution 70/1 – Transforming our world: The 2030 Agenda for Sustainable Development*, New York.



**Prof. Dr. Jörg Faust**  
Direktor des DEval



**Ida Verspohl**  
Evaluatorin

Das Deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) ist vom Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) mandatiert, Maßnahmen der deutschen Entwicklungszusammenarbeit unabhängig und nachvollziehbar zu analysieren und zu bewerten. Mit seinen Evaluierungen trägt das Institut dazu bei, die Entscheidungsgrundlage für eine wirksame Gestaltung des Politikfeldes zu verbessern und Ergebnisse der Entwicklungszusammenarbeit transparenter zu machen.